Inter- and Intra-Operator Variability in the Reading of Indirect Immunofluorescence Assays for the Serological Diagnosis of Scrub Typhus and Murine Typhus

Rattanaporn Phetsouvanh, Thaksinaporn Thojaikong†, Phonlavanh Phoumin, Bountoy Sibounheuang, Koukeo Phommason, Vilada Chansamouth, Sue J. Lee, Paul N. Newton, and Stuart D. Blacksell*

Lao–Oxford–Mahosot Hospital–Wellcome Trust Research Unit, Microbiology Laboratory, Mahosot Hospital, Vientiane, Lao People’s Democratic Republic; Mahidol–Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand; Centre for Clinical Vaccinology and Tropical Medicine, Nuffield Department of Clinical Medicine, Churchill Hospital, Oxford, England, United Kingdom

Abstract. Inter- and intra-observer variation was examined among six microbiologists who read 50 scrub typhus (ST) and murine typhus (MT) indirect immunofluorescence assay (IFA) immunoglobulin M (IgM) slides. Inter-observer agreement was moderate (κ = 0.45) for MT and fair (κ = 0.32) for ST, and was significantly correlated with experience (P = 0.03 and P = 0.044, respectively). χ-scores for intra-observer agreement between morning and afternoon readings (range = 0.35–0.86) were not correlated between years of experience for ST and MT IFAs (Spearman’s r = 0.31, P = 0.54 and P = 0.14, respectively; P = 0.78). Storage at 4°C for 2 days showed a change from positive to negative in 20–32% of slides. Although the titers did not dramatically change after 14 days of storage, the final interpretation (positive to negative) did change in 36–50% of samples, and it, therefore, recommended that slides should be read as soon as possible after processing.

INTRODUCTION

Rickettsial diseases infecting humans are caused by obligate intracellular organisms of the genera Rickettsia and Orientia transmitted by arthropod vectors such as ticks, fleas, lice, and mites. The most common clinical manifestations are fever, headache, and myalgia, making clinical diagnosis difficult because of similarities to other undifferentiated fevers, such as malaria, dengue, and leptospirosis.1 The laboratory diagnosis of rickettsioses conventionally relies on serological tests, such as enzyme-linked immunosorbent assay (ELISA) and indirect microimmunofluorescence assay (IFA), and less tests, such as enzyme-linked immunosorbent assay (ELISA) diagnosis of rickettsioses conventionally relies on serological (or greater) rise in titer between paired sera.3 This interpretation serum without clear justification, rather than a fourfold increase in titer.8

Study site and patient samples. This study was performed on sera from patients admitted to the Adult Infectious Dis-

MATERIALS AND METHODS

Study site and patient samples. This study was performed on sera from patients admitted to the Adult Infectious Dis-

*Address correspondence to Stuart D. Blacksell, Faculty of Tropical Medicine, Mahidol University, 420/6 Rajvithi Rd., Bangkok 10400, Thailand. E-mail: stuart@tropmedres.ac

†Deceased.
Intra-observer variation. Each of the six operators read the 50 scrub typhus IgM slides and 50 murine typhus IgM slides in the morning and afternoon of the same day. The slides were relabeled by an independent investigator between the morning and afternoon sessions to blind the operators to the identity of the slides. A \( \kappa \)-statistic was calculated for each operator’s 50 paired readings for scrub typhus and 50 paired readings for murine typhus with a weighting matrix described below.

**Determination of the effects of 4°C IFA slide storage.** To determine the effect of slide storage at \(-4^\circ\text{C}\) on IFA results, scrub typhus and murine typhus IgM and IgG slides were prepared from the admission and convalescence sera of 50 patients (i.e., 200 slides total). These patients were the same patients as described for the observer variation study. The slides were read immediately after processing (day 0 [D0]), stored in a refrigerator at \(-4^\circ\text{C}\), and reread after 2 (D2) and 14 days (D14) of storage by the same six microscopists as above. At each reading time point, the slides were relabeled and randomized by a person who was not involved in slide reading to blind the operators to slide identity.

**Statistical analysis.** All analyses were performed using STATA version 10 (Stata Corp., College Station, TX). Agreement among the six operators was calculated using the \( \kappa \)-statistic, which was interpreted as follows\(^5\): \( \kappa \leq 0.20 \) (poor), \( 0.21 \leq \kappa \leq 0.40 \) (fair), \( 0.41 \leq \kappa \leq 0.60 \) (moderate), \( 0.61 \leq \kappa \leq 0.80 \) (good), and \( \kappa > 0.80 \) (very good).

For the \( \kappa \)-value calculation, a weighting matrix was used, where one indicated perfect agreement (when paired readings were assigned the same titer) and a weight of 0.60 meant two-thirds agreement (used when paired readings differed by only \( \pm \) one titer). All other paired readings were in complete disagreement (i.e., paired readings differed by \( \pm \) two or more titers) and given a weight of zero.

To determine whether there was any correlation between \( \kappa \)-score and years of experience, each pair-wise combination of the six microscopists (i.e., 15 pairs) was ranked into four groups (i.e., inexperenced, trainee, experienced, and expert). Any pair that included the operator with less than 1 year of experience was considered inexperenced, and the microscopists E and F pair had the highest expert rank. Any pair that included the operator with 1 year of experience was considered experienced, and all other pairs were ranked experienced; \( \kappa \)-scores were calculated for each pair-wise combination of the six microscopists, and these scores were compared against the experience ranking using a Spearman’s rank correlation coefficient.

**RESULTS**

**IgM IFA results read by an expert microscopist.** The results from the microscopist with most experience (10 years; expert microscopist) were used to give the true status of the samples. In summary, for murine typhus IgM IFA results, 25 (50%) slides had titers of \( \leq 1:400 \) titer, and 25 (50%) slides had titers of \( > 1:400 \) titer (1:400 [4%], 1:800 [16%], 1:1,600 [16%], and \( \geq 1:3,200 \) [14%]); for scrub typhus IgM IFA results, 17 (34%) slides gave titers of \( \leq 1:400 \), 33 (66%) slides gave titers of \( \geq 1:400 \) (1:400 [14%], 1:800 [16%], 1:1,600 [8%], and \( \geq 1:3,200 \) [28%]).

**Inter-operator agreement.** \( \kappa \)-analysis overall agreement between the six microscopists was moderate (\( \kappa = 0.45 \)) for murine typhus IgM IFA titers and fair (\( \kappa = 0.32 \)) for scrub typhus IgM IFA titers (Table 1). The range of \( \kappa \)-scores for

---

**Table 2**

<table>
<thead>
<tr>
<th>Microscopist</th>
<th>Years of experience</th>
<th>Experience status</th>
<th>Percentage agreement</th>
<th>( \kappa )-value</th>
<th>Spearman’s ( \rho ) (P value)</th>
<th>Microscopist</th>
<th>Years of experience</th>
<th>Experience status</th>
<th>Percentage agreement</th>
<th>( \kappa )-value</th>
<th>Spearman’s ( \rho ) (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.1</td>
<td>Inexperienced</td>
<td>88.0</td>
<td>0.76</td>
<td>0.14 (P = 0.78)</td>
<td>B</td>
<td>1</td>
<td>Trainee</td>
<td>71.6</td>
<td>0.35</td>
<td>0.62 (P = 0.54)</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>Trainee</td>
<td>92.4</td>
<td>0.86</td>
<td>1</td>
<td>D</td>
<td>3</td>
<td>Experienced</td>
<td>82.8</td>
<td>0.67</td>
<td>0.62 (P = 0.54)</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>Experienced</td>
<td>92.4</td>
<td>0.86</td>
<td></td>
<td>E</td>
<td>5</td>
<td>Expert</td>
<td>92.4</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>Experienced</td>
<td>88.8</td>
<td>0.71</td>
<td></td>
<td>F</td>
<td>10</td>
<td>Expert</td>
<td>88.8</td>
<td>0.71</td>
<td></td>
</tr>
</tbody>
</table>

---
Murine typhus IgM IFA titers was 0.39 (microscopists A and D) and 0.76 (microscopists B and C), and the range of k-scores for scrub typhus IgM IFAs was 0.20 (microscopists A and D) and 0.75 (microscopists D and E). Years of experience and intra-operator k-score were significantly correlated for scrub typhus (Spearman’sρ = 0.70, P = 0.004) and murine typhus (Spearman’sρ = 0.55, P = 0.03).

Intra-operator variation. Overall, the k-values and percent agreement between morning and afternoon readings for each operator were high (Table 2), with k-values indicating fair to very good agreement (0.35–0.86) for the murine typhus slides. For scrub typhus slides, agreement between the morning and afternoon readings ranged from moderate to very good (0.52–0.86) for all six microscopists. No correlation between years of experience and intra-operator k-score was evident for scrub typhus (Spearman’sρ = 0.31, P = 0.54) or murine typhus (Spearman’sρ = 0.14, P = 0.78).

Effect of storage at −4°C. The slides were stored in a refrigerator for 14 days with a mean daily temperature of 4.7°C (95% confidence interval = 3.8–5.6; minimum = 1.4°C, maximum = 9.2°C). IgM and IgG antibody titers for both murine and scrub typhus IFAs showed a general reduction with increasing storage time (Tables 3–5). Murine typhus IFA results tended to be more stable (no change in titer; range = 56–80% of readings) (Table 4) than scrub typhus IFA results (36–64% of readings) (Table 3). Total percentage of IFA titer decreases at day 2 of storage between admission and convalescent samples was reasonably consistent for scrub typhus IgM (30% and 32%, for admission and convalescent sera, respectively) and IgG (both 40%) and murine typhus IgM (28% and 30%) and IgG (28 and 14%). Decreases between admission and convalescent samples were noted at day 14 of storage for scrub typhus IgM (38% and 32%) and IgG (64% and 38%) and murine typhus IgM (30% and 28%) and IgG (34% and 16%). Small increases in titers were also noted, with the largest increase being for scrub typhus IgM admission and murine typhus IgM convalescent titers. Despite the reasonably stable overall percentage change between 2 and 14 days of storage, the magnitude of the change did increase with storage duration (most notably with scrub typhus IFA titers and to a lesser extent, with murine typhus IFA titers), which affected the final interpretation of the result. Storage at 4°C for 2 days showed a change from positive to negative in 20% (murine typhus IgG and scrub typhus IgM) to 32% (murine typhus IgM) of slides (Table 5), and whereas the titers did not dramatically change after 14 days of storage, the final interpretation (positive to negative) did change in 36% (murine typhus IgM/IgG and scrub typhus IgM) to 50% (scrub typhus IgG) of samples.

DISCUSSION

IFA interpretation requires microscopists to use their judgment to determine the fluorescence endpoint, and therefore,
it is inherently subjective. In this study, we found that overall inter-microscopist agreement using κ-scores was fair (scrub typhus IgM IFA) to moderate (murine typhus IgM IFA) (Table 1), with a κ-score range of 0.20 (poor) to 0.76 (good). The results cast doubt on the comparability of results between different studies. Furthermore, they highlight the need for using as few microscopists as possible when reading IFAs titers in a large study. Similar subjectivity has been noted in other studies requiring the interpretations of individual operators. In a study of inter-observer and intra-observer variability among pathologists in lymph node assessments, 15 slides were assessed by 10 pathologists, and significant disagreement in the size of the smallest countable node was noted. An analysis of thin-preparation Papanicolaou tests among 19 cytotechnologists from three different laboratories also showed significant variability.

Microscopist inexperience plays a role in the lack of agreement between IFA microscopists, with significant correlation between κ-scores and years of experience for both scrub and murine typhus IFAs. The poorest agreement was between the least experienced operator (microscopist A) and the other readers; however, generally higher agreement was noted between all other microscopists of various levels of experience. These results suggest that IFA microscopists need a minimum of 12 months training before their IFA results may be considered reliable or that they must buddy with other trainee or experienced microscopists during a period of probation.

Intra-operator κ-scores and agreement between morning and afternoon IFA readings showed variation, but it was generally less than the variation for inter-operator comparison with the same samples. Nevertheless, it is concerning that reproducibility of the titers cannot be guaranteed between two separate readings on the same day. Interestingly, the intra-operator results were not significantly influenced by the experience of the reader, with even the inexperienced reader having good (0.62) to very good (0.76) κ-scores.

Overall κ-scores were generally higher for murine typhus IFA slide reading than scores for scrub typhus IFA slide reading. These results were unexpected, because murine typhus IFA slides have previously been considered (Blacksell SD, unpublished results) to be more difficult to read; the extracellular nature of the R. typhi organism on the IFA slides has a fine gold dust-like appearance and may be confused with artifacts. Scrub typhus IFA slides, however, tend to be simpler for the inexperienced microscopist to interpret, because the O. tsutsugamushi fluorescence is intercellular and more defined. One explanation may be that microscopists spend more time reading the murine typhus slides because of their perceived difficulty of assessment, and this investment of time gives a more reliable result. Antibody titers can influence the κ-scores. The dilution series in this study varied from low titers (<1:400) to high titers (≥1:3,200). Negative sera or sera approaching the positive/negative threshold give more variable results because of the subjective nature of reading. This subjectivity may have contributed to the low κ-scores because of the relatively high number of negative and low to moderate number of positive samples examined.

This study also showed that IFA slides should be read as soon as possible after processing. A change in final diagnostic interpretation of scrub and murine typhus IgM/IgG titers was noted for admission and convalescent sample IFAs stored for 2 days at 4°C, and although the titers did not dramatically change at 14 days of storage at 4°C, the final interpretation (from positive to negative) using the Coleman criteria changed for 30–48% of samples. Interestingly, up to 8% of the samples also showed increases in titers during the storage period, presumably because of intra-operator variation.

In conclusion, the subjective nature of reading IFA slides for titer endpoints, which is compounded by issues of inter- and intra-microscopist variation, microscopist experience, interpretation difficulties when approaching the diagnostic threshold, and storage issues (especially when large numbers of samples are processed), makes the diagnosis of rickettsial illness using IFA variable at best and at worst, unreliable. It is imperative that alternative objective and reliable means of serological gold standard diagnosis are developed and validated without delay. The results also emphasize the importance of internal and external quality assurance schemes for rickettsial IFA slide reading.

Revised May 20, 2012. Accepted for publication September 11, 2012.
Published online March 11, 2013.

Acknowledgments: We are grateful to all the patients who participated in this study, the staff of the Microbiology Laboratory, and the ward doctors and nurses. We are grateful to His Excellency Dr. Ponmek Dalaloy and Professor Sommone Phounsavath, Director of Mahosot Hospital, for their support for this study, which was part of the Lao–Oxford–Mahosot Hospital–Wellcome Trust Research Unit funded by the Wellcome Trust of Great Britain.

Authors’ addresses: Rattanaphone Phetsouvanh, Lao–Oxford–Mahosot Hospital–Wellcome Trust Research Unit Mahosot Hospital, Vientiane, Lao People’s Democratic Republic, E-mail: rattanaphone@tropmedres.ac, Phonnalavanh Phoumin, Bounyot Sibounheuang, Koukeo Phommasone, Vilada Chansamouth, and Paul N. Newton, Lao–Oxford–Mahosot Hospital–Wellcome Trust Research Unit, Microbiology Laboratory, Mahosot Hospital, Vientiane, Lao People’s Democratic Republic, E-mail: phonealavanh@tropmedres.ac, bountoy@tropmedres.ac, koukeo@tropmedres.ac, vilada@tropmedres.ac, and paul@tropmedres.ac, Sue J. Lee, Mahidol–Oxford Tropical Medicine Research Unit, Bangkok, Thailand, E-mail: sue@tropmedres.ac, Stuart D. Blacksell, Mahidol–Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, E-mail: stuart@tropmedres.ac.

Reprint requests: Stuart D. Blacksell, Mahidol–Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol

Table 5

<table>
<thead>
<tr>
<th>Combined admission and convalescent</th>
<th>N = 100</th>
<th>Murine typhus (N = 50)</th>
<th>Scrub typhus (N = 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IgM</td>
<td>IgG</td>
</tr>
<tr>
<td>Day 2</td>
<td>Percent change to positive</td>
<td>7 (14)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Day 2</td>
<td>Percent change to negative</td>
<td>16 (32)</td>
<td>10 (20)</td>
</tr>
<tr>
<td>Day 14</td>
<td>Percent change to positive</td>
<td>1 (2)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Day 14</td>
<td>Percent change to negative</td>
<td>18 (36)</td>
<td>18 (36)</td>
</tr>
</tbody>
</table>
REFERENCES